

Ch 2, Data Analytics: Drawing Conclusions, Part 1

Data Sources, p 91-107

1. Why is it important to know the source of data and its “chain of custody”?
It is important to know the provenance of data, as you would have blind trust in data because you cannot contact the data's creators.
2. List some easily accessible data sites.
Wikipedia, Trove, Atlas of Living Australia, CAIRO, ABS, Data.gov.au, Melbourne data
3. What are the advantages of using different data acquisition techniques?
These sites provide different opinions, or give availability to cross-referencing data.

Acquiring Data, p 94

4. Distinguish between open & closed questions. List some strengths and weaknesses of both.
Open Questions: Often used in interviews, and do not limit an answer, there are no pre-determined set of options nor a limit on length, type or content.
Strength: Allows for a higher-detailed response, allows freedom of response from the participant, gives more accurate and elaborate results.
Weakness: Takes a long time to process and categorise.
Closed Questions: Usually found in surveys, exams and forms. They force you to answer in a particular way because only allow pre-determined responses, such as yes or no, or a checkbox.
Strength: Quick processing, more questions can be asked in shorter time, easy to process.
Weakness: Not detailed enough, may not cover the participants' response, so they pick the closest answer rather than their actual answer.

Faults in questions, p 96

5. List some types of questions that are not effective.
“Did you see the accused come out of the house at 2 a.m. on the fifth of May this year?” “Was the accused wearing a red jumper today?”, “Were you at the football field on Thursday playing football?”

Conducting interviews, p 97

6. List some strengths and weaknesses with interviews as a technique for acquiring data.
Strength: Questions can be tailored to your liking, and can pursue a line of questioning to elicit more information upon receiving an interesting answer. You can also pick up on the interviewee's tone of voice – interest, disinterest, anger, happiness etc. Body language can also be spotted, as well as facial expressions, pauses, hesitancy, energy, lethargy and other non-verbal elements of communication.
Weakness: Face-to-face interviews can be time-consuming and expensive. Sometimes, the interviewee may tend to say the ‘right’ thing in order to please the interviewer, or to appeal to their public reputation rather than give the correct answer or avoid embarrassment.

Observing subjects, p 98

7. What are the strengths and weaknesses of observation?
Strength: Discloses the truth based on generally ‘right’ answers, such as hospitals asking if patients and visitors have washed their hands (generally, 100% say yes), which can be debunked if observed. It allows for non-verbal communication methods as in interviews. As subjects may not be entirely capable of reporting upon themselves with clarity, and tend to be biased as they are reporting on themselves, you may be able to see where they are wrong and where they are lying. Covert (unseen) observation is more accurate.

Weakness: Covert observation may be unethical or illegal. Observation is incredibly time-consuming, and can take a great deal of money and labour. It only indicates what subjects do, and not why they do it. Subjects also tend to perform better if they know that they are being observed, which can affect realistic results. Interpretation of questions may be different for different people, or some may be biased towards the observation question.

Surveys and questionnaires, p 99

8. What are the advantages and disadvantages of these techniques?

Advantages: If observation is overt, trust, consent and confidence from subjects can be gained by explaining the purpose of the observation – the same works if the subject is reassured that they will not be punished or judged. If records are taken, they can always be read over – memory may fail. Recording with evaluation rather than interpretation (quotes, events etc) provide the most accurate results. Recording field notes during the observation (which are open ended recordings of things that were observed which weren't anticipated when research began), as they add detail

Disadvantages: Field-notes do not answer pre-planned questions. Observing may distort the natural behaviour of subjects, as interference may throw them off. Observer must remain professionally detached, as emotional attachment may distort an observer's objective judgement.

Selecting representative samples, p 104

9. What types of selection bias exist?

Wealth-based bias exists – for example; the American Literacy Digest in 1936 created a survey depending on which president would win the election. They found that Alfred Landon would win indefinitely, but Roosevelt won by a long shot. This was because of the economic bias, as only middle and upper class citizens voted, which tended to be more Republican. Another bias is timing – some subjects may not have the time to answer the question, and will either throw an answer for the sake of answering, or refuse to answer entirely. Another type is self-selection bias – A pollster in a supermarket found that an unusually high number of shoppers did charity work. This was perhaps because the sort of person who does charity work is also the sort of kind, empathetic person who would stop to help a pollster in a supermarket. Just choosing to participate in a survey self-selects a certain type of respondent which can lead to biased results

Querying resources, p 105

10. What is the role of built-in services in databases?

Built-in services in databases are used to assist in collecting data that is relevant to the hypothesis.

11. Give examples of 2 types of queries.

Query By Example (QBE) uses a database's graphical user interface (GUI) to let a user type the data they wish to find in a field to be searched. On the other hand, Structured Query Language (SQL) is the standard method of requesting data from databases, and is also the textual equivalent of QBE.

12. What is an application programming interface, (API)?

An application of an API is to let programmers create a gateway so other programs can interact with their software.

13. Why should the sources of data you use be acknowledged?

The sources of data you use should be acknowledged so that it is clear that you are not claiming another author's work as your own, so that readers can find the original source to get more information, moral rights are observed, and copyright 'fair use' legalities are observed.